DOCUMENT RESUME

ED 209 348 TM 810 903

AUTHOR Baldauf, Richard B., Jr.; Bisazza, John A.
TITLE The Reliability and Interrelationships of Three Forms of the Michigan Test of Aural Comprehension.
PUB DATE [Nov 81]
NOTE 10p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *English (Second Language); Examiners; Experimenter Characteristics; Higher Education; *Listening Comprehension Tests; *Test Reliability
IDENTIFIERS *Michigan Test of Aural Comprehension

ABSTRACT

One hundred eighty limited-English-speaking college students from the Philippines, Hong Kong, India, Japan, Thailand, and Taiwan were administered the Michigan Test of Aural Comprehension (MTAC). An analysis of the MTAC indicated that the three unrestricted forms of the test show satisfactory internal consistency reliability comparable to that found for other listening comprehension tests. However, the results strongly suggest that counselors should not treat the three forms as being statistically equivalent. Furthermore, there is strong evidence that speaker variation, contrary to the test manuals' assertions, can effect test results. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED209348

THE RELIABILITY AND INTERRELATIONSHIPS OF THREE FORMS OF THE MICHIGAN TEST OF AURAL COMPREHENSION

Richard B. Baldauf, Jr.

John A. Bisazza

East-West Culture Learning Institute

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ABSTRACT

An examination of test results for ESL and EFL college students from six countries for Forms A, B, and C of the Michigan Test of Aural Comprehension provided test reliability estimates previously unavailable. Further analysis indicated a lack of equivalence between forms and the need for caution in selecting test administrators with different English language accents.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R.B. Baldauf, Jr.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 503

RUNNING HEADING: MTAC Three Forms

THE RELIABILITY AND INTERRELATIONSHIPS OF THREE FORMS OF THE MICHIGAN TEST OF AURAL COMPREHENSION

Oral comprehension is an essential element in the repertoire of skills necessary for the non-native-English-speaking applicant who wants to study at an institution of higher education in the United States (Smith & Bisazza, submitted).

A glance at Buros (1978) indicates that there are few non-secure tests the counselor can use to evaluate initial or progressive development of aural comprehension skills of such students. One test which is available is the Michigan Test of Aural Comprehension (MTAC) (Lado, 1957).

The MTAC is a 40-minute test designed to measure 'understanding of spoken English by persons whose native language is not English' (Lado, 1957, p.1). The test seems suitable for evaluation studies because three parallel non-secure forms are available. However, there is little empirical evidence in the test manual or in the literature about the reliability of the MTAC, or the interrelationship of the three forms of the test. John B. Carroll in a review of the tests noted that the manual 'claims that the test scores from the three forms are equivalent, but gives no information concerning equivalent test reliabilities or the like' (Buros, 1978, p.188). This study was undertaken to collect such data using students from six countries, three where English is a second (ESL) and three where English is a foreign (EFL) language.

Methods

Subjects

The subjects, 30 each from the Philippines, Hong Kong, India, Japan, Thailand, and Taiwan, were enrolled in beginning English language university programs. The six groups were roughly balanced for sex and age and all had had at least five years of prior English language study. In the first three countries mentioned, English is spoken as a second language while in the latter three, English is a foreign language.

Instruments

The MTAC is a two part 40-minute examination with three conceptually equivalent forms, each of which consists of 60 sentences and short paragraphs to be read aloud by an examiner. Students respond using a booklet containing three pictures or phrases, each of which is the correct answer to a question for one of the three forms. The study used the three earliest forms of the MTAC since they are not currently restricted and are being used in Asia and the Pacific.

The test manual claims that forms A, B, and C are of approximately equal difficulty. An alternate-forms reliability between forms A and C of .87 and a split-half reliability of .88 for form A are given. No norms are available, but proficiency and progress standards based on percent correct scores are provided.

Procedures

The design of the study uses a complex latin squares design procedure (Bruning and Kintz, 1977). The three forms of the test were

given on separate days within a ten-day period to students in each of the six countries. Three variables, order of presentation, test form, and speaker were balanced within each of the three countries in the ESL and EFL groups. Taped presentations of the three proficient speakers of English with different accents (native, Indian ESL, and Japanese EFL) were used to ensure a standardized administration of the test materials.

Results

Table 1 provides Kuder-Richardson (KR) formula 20 estimates of internal consistency and their corresponding standard errors of estimate by English language status and country for percentage correct scores from forms A, B, and C of the MTAC. Alternate forms reliability estimates based on Pearson r's were calculated between three test forms.

INSERT TABLE 1 HERE

The data were first analyzed to see if there was an order of presentation effect. A significant percentage score increment of 2.5 points was found between both pairs of test administration means. To control for this effect, T-scores were generated for each order of presentation. The resulting data, with order of presentation now controlled, was then analyzed using a complex latin squares design which balanced test form (A, B, C) and type of speaker (native, ESL, EFL), within each language status (ESL, EFL) group. These results are presented in Table 2.

INSERT TABLE 2 HERE

Discussion

As Table 1 indicates, the internal consistency estimates found in this study averaged in the mid to upper .70's with a standard error of about 4.50. While these values are lower than the split-half reliability of .88 with a standard error of .02 (surely this latter figure is incorrect) reported in the manual, they are consistent with the somewhat lower reliabilities generally associated with listening comprehension tests. The standard error of estimate of about 4.50 for the percentage scores should provide test users with a more accurate estimate for evaluating student score differences. As the following analysis indicates, the lower median test-retest reliabilities of .54, .57, and .64 may be due in part to the confounding effects created by the design of the study when significant variables effects are found.

The latin squares analysis (Table 2) indicates that the three forms of the test are not statistically equivalent ($F_2,348 = 47.89$, $p<.001$). Scores on form A averaged one quarter of a standard deviation below those for form C, while scores on form B were nearly half a standard deviation below the scores on form C. Furthermore, the significant form by type interaction ($F_2,348 = 35.18$, $p<.001$) indicates that these differences were not consistant for ESL and EFL students. On form A ESL student scores averaged 1.3 standard deviations above EFL students. Student scores for form B were only 1.0 standard deviation apart while scores on form C were 0.55 standard

deviations apart. These results strongly suggest that alternate forms pre-post test designs should be avoided, especially with heterogeneous groups.

The manual states that 'Any teacher with a reasonably clear voice who speaks a variety of Standard English can give the test.' (Lado, 1957, p.2). However, the speaker effect found in this study (F_2,348 = 102.16, $p<.001$) strongly suggests this is not the case. Students averaged 0.21 standard deviations better on the test when it was given by a native speaker than when it was given by a fluent EFL speaker from Japan. When the tests were given by a fluent ESL speaker from India, they averaged 0.64 of a standard deviation below those results obtained by the native speaker. Speaker variation clearly effects test results, at least when the presentation is made using tape recorded presentations. Test designs should employ the same or a comparable speaker to avoid biasing test results.

Conclusions

An analysis of the MTAC indicated that the three unrestricted forms of the test show satisfactory internal consistency reliability comparable to that found for other listening comprehension tests. However, the results strongly suggest that counselors should not treat the three forms as being statistically equivalent. Furthermore, there is strong evidence that speaker variation, contrary to the test manuals' assertions, can effect test results.

REFERENCES

Bruning, J.L. and Kintz, B.L. *Computational Handbook of Statistics* (2nd). Glenview, Ill.: Scott, Foresman, 1977.

Buros, O.K. *The Eighth Mental Measurements Yearbook.* Hyland Park, N.J.: Graphon Press, 1978.

Lado, R. *Test of Aural Comprehension: Description and Norms.* English Language Institute, Ann Arbor: University of Michigan, 1957.

Smith, L. and Bisazza, J.A. The comprehensibility of 3 varieties of English for speakers in 7 countries. Submitted for publication.

TABLE 1

Internal Consistency, Standard Errors, and Alternate Forms Reliability Estimates for the Michigan Test of Aural Comprehension by Language Status and Country

Language Status / Country	N	KR_{21} Internal Consistency						Alternate Forms		
		r_{AA}	SE_M	r_{BB}	SE_M	r_{CC}	SE_M	r_{AB}	r_{AC}	r_{BC}
ESL Students										
Hong Kong	30	.78	3.80	.81	3.42	.74	4.32	.70	.48	.33
India	30	.92	4.32	.91	4.55	.92	4.11	.77	.85	.74
Philippines	30	.78	2.68	.61	4.11	.46	2.83	.34	.17	.18
EFL Students										
Japan	30	.78	4.69	.62	4.84	.64	4.16	.54	.49	.18
Taiwan	30	.59	4.54	.76	4.35	.78	4.56	.57	.65	.75
Thailand	30	.68	4.93	.75	4.90	.76	4.48	.73	.74	.71
Median Values		.78	4.43	.76	4.45	.75	4.24	.64	.57	.54

TABLE 2

Complex Latin Squares ANOVA for Test Form by Speaker by Student Type

Source	SS	df	MS	F
Between Subjects	40,940.00	179		
Student Type	12,231.30	1	12,231.30	10[illegible].17***
Speaker X Form$_b$[1]	7,059.54	2	3,529.77	30.64***
Spk X Form$_b$ X Type$_b$[1]	1,604.16	2	802.08	6.96**
Error$_b$	20,045.00	174	115.20	
Within Subjects	13,524.98	360		
Speaker	3,827.41	2	1,913.70	102.16***
Form	1,794.14	2	897.07	47.89***
Speaker X Form$_w$	0.24	2	0.12	0.01
Speaker X Type	61.23	2	30.61	1.63
Form X Type	1,318.27	2	659.14	35.18***
Spk X Form$_x$ X Type$_w$	4.38	2	2.69	0.14
Error$_w$	6,519.31	348	18.73	
Total	54,464.98	539		

[1]Those sources with mixed between and within components have been tested at the alpha/2 percent level.

** p<.01
*** p<.001